UNIVERSITA' DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI "M. FANNO"

CORSO DI LAUREA MAGISTRALE IN
Economics and Finance

TESI DI LAUREA

"Rounding patterns of survival probabilistic expectations and their consequences on inference."

RELATORE:

CH.MA PROF.SSA: Dal Bianco Chiara

LAUREANDO: Cicero Simone

MATRICOLA N. 2062942

ANNO ACCADEMICO 2022 – 2023

Dichiaro di aver preso visione del "Regolamento antiplagio" approvato dal Consiglio del Dipartimento di Scienze Economiche e Aziendali e, consapevole delle conseguenze derivanti da dichiarazioni mendaci, dichiaro che il presente lavoro non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere. Dichiaro inoltre che tutte le fonti utilizzate per la realizzazione del presente lavoro, inclusi i materiali digitali, sono state correttamente citate nel corpo del testo e nella sezione 'Riferimenti bibliografici'.

I hereby declare that I have read and understood the "Anti-plagiarism rules and regulations" approved by the Council of the Department of Economics and Management and I am aware of the consequences of making false statements. I declare that this piece of work has not been previously submitted – either fully or partially – for fulfilling the requirements of an academic degree, whether in Italy or abroad. Furthermore, I declare that the references used for this work – including the digital materials – have been appropriately cited and acknowledged in the text and in the section 'References'.

Firma (signature) ….Simone Cicero……

INDEX

I ABSTRACT

The focus of this thesis is on subjective expectations, how they are formed and their consequences in the financial market participation.

Data on subjective expectations are taken from the SHARE survey: the Survey of Health, Aging and Retirement in Europe, which will also allow to make comparisons of results across countries. The thesis analyze the measurement error issue when using subjective expectations, due in particular to the fact that people tend to "round up" their responses by often using easily communicable values.

I estimate the magnitude of this measurement error through an algorithm that transforms individuals' point responses into interval responses based on individual response patterns.

In the final part of the thesis, I present an empirical application in which I analyze the relationship between investment in the financial market and the probability with which the individual believes he will survive beyond a certain target age.

I compare model estimates using point responses and interval responses in turns.

The results suggest that measurement error in this specific empirical application has no significant impact on the conclusions.

II ACKNOWLEDGEMENTS

This paper uses data from SHARE Waves 1, 2, 3, 4, 5, 6, 7, 8 and 9 (DOIs: 10.6103/SHARE.w1.800, 10.6103/SHARE.w2.800, 10.6103/SHARE.w3.800, 10.6103/SHARE.w4.800, 10.6103/SHARE.w5.800, 10.6103/SHARE.w6.800, 10.6103/SHARE.w7.800, 10.6103/SHARE.w8.800, 10.6103/SHARE.w8ca.800, 10.6103/SHARE.w9ca800) see Börsch-Supan et al. (2013) for methodological details.(1) The SHARE data collection has been funded by the European Commission, DG RTD through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812), FP7 (SHARE-PREP: GA N°211909, SHARE-LEAP: GA N°227822, SHARE M4: GA N°261982, DASISH: GA N°283646) and Horizon 2020 (SHARE-DEV3: GA N°676536, SHARE-COHESION: GA N°870628, SERISS: GA N°654221, SSHOC: GA N°823782, SHARE-COVID19: GA N°101015924) and by DG Employment, Social Affairs & Inclusion through VS 2015/0195, VS 2016/0135, VS 2018/0285, VS 2019/0332, and VS 2020/0313. Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, HHSN271201300071C, RAG052527A) and from various national funding sources is gratefully acknowledged (see www.share-project.org).

III INTRODUCTION

The focus of this thesis is on individual expectations about future events and their use in economic analysis.

The formation of expectations has always been a relevant topic, as early as 80 years ago, Lachman (1943) said in his paper that expectations are an essential element of economics. He argued that individuals' expectations influence their economic decisions, and that these decisions, in turn, influence the economy as a whole. He defined expectations as an individual's beliefs about future events and argues that expectations can be based on concrete information or vague and unpredictable. Lachman concluded by saying that economists need to understand how individuals form expectations and how expectations influence economic behavior in order to formulate more effective economic policies.

Until now, to predict how economic agents form their expectations, many economists have relied on revealed preference theory: a theory dating back to the 1950s which says that by observing an individual's choices we can learn about his preferences and consequently prove and predict his decisions or expectations.

This theory has two main assumptions: the first one is that economic agents are rational: according to this theory, all this would be possible because the rational individual is able to process well all the information he has before making a choice and he is able to maximize his utility through its choices. The second important assumption of this theory is that all individuals start with the same set of information when they have to make a choice.

Both assumptions are very strong and strict, and over the years empirical evidence has shown that they are not so realistic.

Behavioral economics represented a real turning point in the study of economic individuals by introducing for the first time the idea that investors are not always rational and that their choices can be influenced by cognitive biases and other factors. Also, the assumption that all individuals have the same information when making a choice is very unrealistic. The information set is very important when making decisions, especially for the fact that it influences not only the choices that affect the present but also those that affect the future as pointed out by Van der Klaauw (2012). He stated that the current information set influences the preferences of each individual and is then likely to determine both current saving behavior and future social security benefits. Thus, he concluded that treating expectations as exogenous explanatory variables is likely to lead to endogeneity biases.

As we have seen, therefore, neoclassical theory cannot be taken as a starting point for creating a model that tries to study the formation of expectations of economic individuals.

An innovative approach to predict choice behavior has been proposed by Mansky (2004) and involves combining choice data of each individual with other data. Specifically, he refers to subjective expectations that can be used as a starting point to relax or validate assumptions about expectations.

Manski's proposed innovation lies in using survey data to try to predict the preferences and thus the choices of economic agents.

The biggest problem with the use of subjective expectations data is the measurement error, which is the failure to measure the right "value" in the case of survey data because of bias.

Most often, the measurement error of a given variable is assumed to be independent of the true level of that and all other variables in the model, the measurement error of other variables, and stochastic disturbance.

However, Bound, Brown and Mathiowez (2001) explained that measurement error in survey data cannot be considered as the "classical" measurement error and explain why. In particular, the paper explains that cognitive processes are involved in survey data that could lead the individual to make mistakes, as well as social desirability, that is, the desire to give an answer that is socially acceptable. They further conclude by saying that survey conditions (the way questions are asked, the way they are collected, the interviewer's characteristics etc.) can also be a cause of measurement error in survey data.

In practice, the measurement error present in survey data depends on the respondents and it is dependent on the “true” value. Individuals tend to have cognitive biases when answering probability questions and often do not indicate the true value they have in mind, but they find easy to communicate rounded values, such as multiples of 5 and 10, or they might express their lack of understanding of the question by providing answers such as 0,50 or 100.

Several solutions have been proposed to account for measurement error in the expectation questions (see section II for a literature review on the topic), and in this work I follow the approach proposed by by Giustinelli, Manski, and Molinari (2022) consisting of "rounding" the point responses given by each respondent with interval responses. These intervals are constructed using an algorithm that studies the response patterns used by each individual and, based on that, identifies a certain rounding rule that determines whether the interval is wider or narrower.

They use data taken from the HRS (Health and retirement Survey), a longitudinal research study conducted in the United States that collects demographic, health and economic data on older people. Specifically, they consider expectation questions involving personal finance, personal health and general economic condition and follow individuals for 7 waves. Their

working assumption is the stability of each individual's response tendencies across questions and waves.

In this thesis I analyze survey data responses to expectation questions, and in particular I use data from the Survey of Health, Ageing and Retirement in Europe (SHARE). I borrow from Giustinelli et al (2022) and develop an algorithm to round the point responses into intervals, a strategy to account for measurement error included in expectation questions. I use survey responses and the generated intervals in turn to estimate a model that analyzes the correlation between subjective survival probability and financial market participation.

To account for measurement error (interval responses) I impute 10 possible values of the subjective survival probability drawing randomly from the interval. I then apply techniques used for multiple imputations (Little and Rubin, 1987) to estimate the model using these 10 imputations.

Comparing estimation results, I see that in most of the variables considered within the regression the estimates obtained using interval responses give more or less the same value but with a higher standard error.

For what concerns the variable of interest, subjective survival probabilities, the larger standard errors do not imply a change in the main conclusions: a 25 percentage point increase in the subjective probability of survival increases financial market participation by values that range from 1 to 2 percentage points.

This thesis has allowed me to understand that survey data are a great resource to use because they allow to analyze the subjective expectations of each individual and to match them with other subjective characteristics that are very important to better understand how the individuals form their expectations (education, health, wealth, job situation).

Of course, these data also contain measurement error, but in the specific empirical application considered it does not affect in an important way the estimation results.

IV LITERATURE REVIEW

In this section I present a literature review on the use of expectations data in economics and the most relevant issues related to it. I then focus on the empirical application and review the literature on the role of lifespan and survival probability on economic decisions.

The first who introduced the idea that measurement of probabilistic expectations might improve with probabilistic survey research was Juster (1966) who in his paper analyzed responses to a binary purchase decision and concluded that it was more effective to ask individuals for their purchase probabilities rather than for their buying intentions and in particular he proposed questions that associate verbal expression with numerical probabilities.

Following Juster, Morrison (1979) added another important piece of literature regarding the use of expectations. Specifically, he showed in his paper that an individual's expectations play a very important role in his future decisions and that it is important to understand how individuals form their expectations to predict their actions. He came to these conclusions using survey data on purchase intentions for a new car model.

In the first 20 years after Juster's article was published (1966), the topic of expectations and their use did not attract the attention of many economists. Beginning in the 1990s, however, expectations became a central theme in economics and many began to write about them.

In particular, since the 1990s, surveys that use individuals' subjective probabilities as a starting point for studying the expectations of economic agents have begun to become increasingly common. Some of the best known surveys that started in the 1990s and are still being submitted to individuals today are the Health and Retirement Study (HRS) started in 1995, the Bank of Italy's Survey of Household Income and Wealth (1992) and the Survey of Economic Expectations (1997). Also in other surveys, such as the Survey of Consumers, subjective probability questions have been introduced.

Of course, some economists have pointed out critical issues regarding the use of expectations. One of the best known was Machlup (1946) who explained in his paper that there are two types of expectations: rational and unrealistic. Unrealistic expectations, according to the author, are much more common than people think, and using them can lead to erroneous conclusions.

Later papers pointed out critical issues with surveys collecting data on expectations. In particular, Lichtenstein and Newman (1967) pointed out that survey data can be wrong and identify 3 biases: response bias, selection bias, and measurement bias. They also propose solutions to make survey data more effective: simple questionnaire design, homogenous data collection procedure, and "preparation" to the interviewers who will support the survey.

From that moment on, the focus of economists was on how to use survey data to predict individuals' expectations and, as mentioned in the introduction, I decided to follow the approach proposed by Mansky (2004) which consists of using subjective probabilities (mixed with other individual data) in order to measure expectations.

There is also the psychological aspect of the respondent to consider and in particular the cognitive processes involved in responding to a survey. McClelland and Bolger (1994) in their paper argue that the main cognitive processes used in answering questions are: memory, evaluation, and social constraint. The authors later explain how all three of these processes can lead to errors in the data.

In addition to Giustinelli et al. (2022), there have been other papers that have tried to provide a solution to the measurement error problem in survey data. In particular, Saris and Revilla (2016) in their paper proposed to take into account additional information about the measurement error, such as estimates of the reliability or validity of the survey questions, in order to estimate the measurement error and correct the data.

Blattman et al. (2016) proposed to use qualitative interviews in addition to quantitative interviews submitted to study participants as a solution to measurement error. During the interviews, participants are asked open-ended questions about their sensitive behaviors. Then, the responses are compared to the responses to the survey questions, and measurement error is estimated and corrected.

In this thesis I use the rounding of punctual responses into interval responses proposed by Giustinelli et al. (2022) to try to account for potential measurement error present in the data I analyze.

Then I use the data obtained to estimate a model that studies the correlation between survival probability and participation in the financial market. For this model I took inspiration from the paper by Spaenjers and Spira (2015). They analyze the relationship between subjective life expectancy and portfolio choice and found that equity portfolio shares are higher for investors with longer horizons.

Not many papers before this one had analyzed the relationship between subjective life expectancy and portfolio choice. In particular, it is important to mention Bloom et al. (2006). In their paper they studied the relationship between subjective life expectancy and saving and consumption decisions. The conclusion was that people with longer life expectancies usually save more and delay retirement than people with lower life expectancies.

After having provided a complete picture of the literature used and the working method that I use during the thesis, in the following section I present the data.

V DATA USED AND DESCRIPTIVE TABLES

For this thesis, the data were taken from the SHARE survey. This survey collects data on health, social, economic and environmental policies on the life trajectories of European citizens. SHARE main focus is on individuals aged 50 years and older who maintain their residence whitin one of the countries involved in SHARE. Individuals who were incarcerated, hospitalized, out of the country for the entire survey duration and individuals who are unable to communicate in the country's language are excluded from the sample.

These data were first collected in 2004 and the first wave of the survey included respondents from 18 different countries. In 2023, the survey involves 28 different countries: Austria, Germany, Sweden, Netherlands, Spain, Italy, France, Denmark, Greece, Switzerland, Belgium, Israel, Czech Republic, Poland, Ireland, Luxembourg, Hungary, Portugal, Estonia, Croatia, Lituania, Bulgaria, Cyprus, Finland, Latvia, Malta, Romania, Slovakia.

Now SHARE counts 8 waves plus two ‘special’ waves submitted during the Corona Virus pandemic crisis (2020 and 2021).

In this thesis I use data from waves 4, 5, and 6, collected in the years 2011, 2013, and 2015, respectively. I chose these three waves because waves 3 and 7 are two peculiar waves since they contain retrospective data (SHARELIFE). In the SHARELIFE waves the focus is on respondents' life history while in all other waves the focus is on current life circumstances. Especially, the choice was also determined by the fact that waves 3 and 7 lack questions about respondents' expectations, which are essential for this thesis.

The countries considered in this paper and for which respondents can be found in all 3 waves are: Austria, Germany, Sweden, Spain, Italy, France, Denmark, Switzerland, Belgium, Czech Republic, Slovenia and Estonia. The fact that there are respondents from various countries within the survey is very interesting because allows to compare response patterns across countries and see if there are response styles that are country specific.

The sample contains 25866 individuals followed in all 3 waves. This is an important aspect. By observing the answers of each individual over time it is possible to understand if there is a "pattern" in his answers to expectations questions, for example if he tends to report some percentages rather than others or to round his answers in specific ways.

The panel analyzed is balanced and contains for each individual responses collected in all the 3 waves. I drop from the sample individuals who were less than 50 years old at the time of the interview (to whom the survey was administered only because they were partners of survey respondents) or more than 80 years old. Individuals younger than 50 are very few and very

old individuals might be inaccurate in reporting expectations due to cognitive limitations related to ageing.

In this thesis I utilize the panel dimension of data for the analysis of response patterns among individuals. This decision is driven by the recognition that to investigate response patterns is required a sufficient number of questions for each individual. Utilizing a single wave of data, which typically contains a limited number of expectation questions, may not provide an adequate dataset to capture the complexity of response patterns effectively. This approach allows for a more extensive set of questions per participant, enabling a more accurate analysis of individual response patterns over time.

Table 1: Descriptive analysis containing mean age, percentage of couples, percentage of women and numerosity, divided by country.

Country	Mean age	(%) couple	(%) women
Austria (7545)	65.36	.688	.575
Germany (2676)	67.14	.778	.529
Sweden (3226)	68.46	.747	.560
Spain (6110)	66.28	.804	.560
Italy (5891)	66.44	.827	.559
France (7504)	64.83	.688	.566
Denmark (4506)	63.84	.779	.532
Switzerland (6705)	64.77	.763	.545
Belgium (8423)	64.29	.708	.555
Czech republic (9002)	65.74	.695	.608
Slovenia	64.92	.751	.578

(4330)			
Estonia (11680)	66.03	.683	.618
Total (77598)	65.52	.731	.573

Source: Börsch-Supan,(2022). Calculation from SHARE waves 4,5 and 6.

Table 1 reports information on the composition of the sample and the main characteristic of the respondents. It is possible to see that the country with the largest number of observations is Estonia while the country with the smallest number of observations is Germany. The average age is similar in all countries, ranging from 63.84 for Denmark to 68.46 for Sweden. Some heterogeneity in average age is due to differences in the timing in which each country joints the survey and on the timing of the inclusion of new respondents (refreshment samples) in the country samples to preserve representativeness of individuals ages 50+.

The variable couple refers to any type of couple (married and unmarried) and in this case the data are quite different between countries because it ranges from 68.25% of people having a partner in Estonia to 82.74% in Italy.

Regarding gender, the lowest percentage among women is registered in Germany (52.99%) and the highest figure is in Estonia (61.78%).

After having examined a general overview divided by countries, I will now take a closer look at the variables that I will use throughout the entire thesis. These variables will also be employed in the final empirical application as control variables.

Table 2: Summary of variables

Variable	Mean	Std. Dev.	Min.	Max.
Age	65.52	7.747	50	80
Women	.5728	.4947	0	1
Single	.2390	.4440	0	1
Children	2.163	1.294	0	17

Isced	2.949	1.424	0	6
Employed	.2662	.4420	0	1
Sphus	3.143	1.046	1	5
Numeracy	4.255	4.253	0	5
Ex 007	48.44	36.32	0	100
Ex 008	48.04	38.61	0	100
Ex 009	64.90	28.68	0	100
Ex 025	46.26	39.21	0	100

Source: Börsch-Supan, (2022). Calculation from SHARE waves 4,5 and 6.

According to table 2, and as can be gauge from Table 1, average age in the sample is 65.52, the 23,9 % of respondents do not have a partner and the 57,3% of respondents are women. Additionally, Table 2 contains information on:

-The number of children, which shows large variability and a relatively high maximum of 17, and tells us that on average SHARE survey respondents have 2 children.

-An indicator of education, specifically the Isced code, which can take values from 0 to 6 where 0 indicates no education and the code 6 indicates bachelor's or equivalent.

-An indicator of the respondents' job status. The employed variable is a dummy variable that takes value 1 if the individual is employed or self-employed and 0 in all other cases. In this case it is possible to see that 26.6 percent of people belong to this category. The majority of individuals (65%), on the other hand, are retired.

-An indicator of health. The acronym Sphus stands for self-perceived health status and can take values from 1 (poor) to 5 (excellent).

- An indicator of numerical ability (numeracy) which can also take values from 0 to 5. Numeracy takes the value of correct answers to a set of questions like:

1. If the chance of getting a disease is 10 percent, how many people out of 1000 (one thousand) would be expected to get the disease? (Possible answers: 10,100,90,900, other)

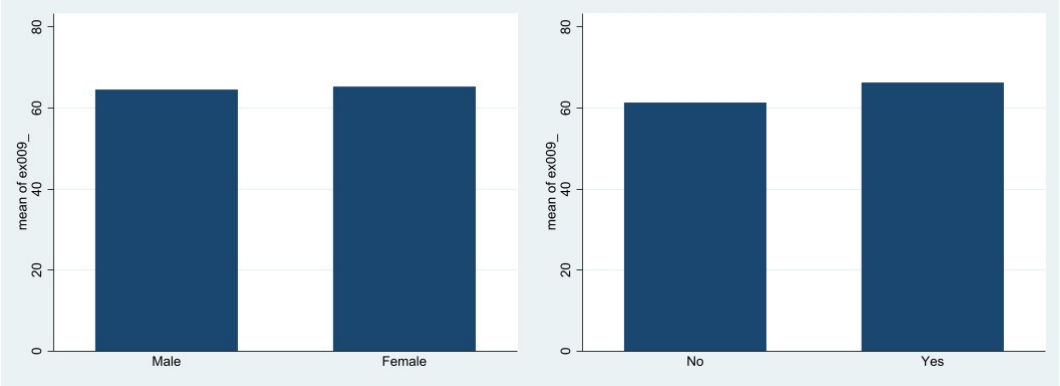
2. In a sale, a shop is selling all items at half price. Before the sale, a sofa costs 300. How much will it cost in the sale? (Possible answers: 150, 600, other)
3. A secondhand car dealer is selling a car for 6,000. This is two-thirds of what it costs new. How much did the car cost new? (Possible answers: 9000, 4000, 8000, 12000, 18000, other)
4. Let's say you have 2000 in a savings account. The account earns ten per cent interest each year. How much would you have in the account at the end of two years? (Possible answers: 2420, 2020, 2040, 2100, 2200, 2400, other).
- An indicator of total income relative to the country. In fact, 4 income quartiles were created for each country.
-Some question about expectations. Specifically:
1. Ex 007: “What are the chances that before you retire the government will reduce the pension which you are entitled to?”
2. Ex 008: “What are the chances that before you retire the government will raise your retirement age?”
3. Ex 009: “What are the chances that you will live to be age <target age> or more?”
4. Ex 025: “Thinking about your work generally and not just your present job, what are the chances that you will be working full-time after you reach age 63?”
Expectation question cover several dimensions: subjective survival probability, future job prospects and future moves the State regarding pension.
The first expectation question analyzed is the one about subjective survival probability. For this part, I refer to question Ex 009 of the SHARE survey, which says: “What are the chances that you will live to be age <target age> or more?”.
The <target age> depends on the age of the respondent at the time of the survey and is defined as follows:

- if age of respondent is <65 the target age is 75.
- if 64 < age of respondent < 70 the target age is 80.
- if 69 < age of respondent < 75 the target age is 85.
- if 74 < age of respondent < 80 the target age is 90.

The first distinction for the responses to question Ex 009 can be made regarding gender and whether the respondent is in a couple. As can be seen from the two graphs 1 and 2, gender does not have a significant impact; in fact, the average response of women (65.19%) is only slightly higher than that of men (64.50%) and this figure is anomalous because on average women live about 5 years longer than men. Being or not being in a couple, on the other hand, has a great impact in respondents' life expectancies, in fact those in a couple gave an average

response of 66.23% compared to 61.33% for people not in a couple. This result was quite predictable as in the literature it is possible to find many papers demonstrating the benefit of living in couples. In particular England (2001) reports that people in couple are happier, healthier, and better off financially and she also gave explanations for this phenomenon: sharing economic and social resources is seen as co-insurance by couples who also benefit from economies of scale, and of course the social factor is also important: being in a couple often allows to connect with other social groups and institutions. In the paper the author talks about the benefits of marriage, but most of the benefits occur in general if you are in a couple with another person.

Graph 1: mean of Ex 009 over gender. Graph 2: mean of Ex 009 over couple.



Source: Börsch-Supan, (2022). Calculation from SHARE waves 4,5 and 6.

From now on, 4 different age ranges are introduced to facilitate the analysis and to consider responses from people who have the same target age in the expectation question about subjective survival probability. The first range goes from 50 to 64 years, the second range goes from 65 to 69 years, the third range goes from 70 to 74 and the last range goes from 75 upwards. This division was made to understand how much age influences survival probability and above all if it is possible to find links between the age of the respondent and one of the characteristics that will be analyzed in the following tables (health, education, work situation and confidence in the financial market).

The analysis shows that age itself, as expected, has a great impact on survival probability: as might be expected the average survival expectation of the first age range (71.08%) is much higher than that of the last range (48.33%), but it is very interesting to look whether survival probabilities changes with age and a set of social-economic characteristics. In particular, the

next tables analyze the correlation (if any) between subjective survival probability and some social, demographic and economic characteristics.

Table 3: Mean answers to question Ex 009 by Isced code and numerosity divided by range.

Isced	(50-64)	(65-69)	(70-74)	(75-80)
0/1	60.74	60.23	54.31	46.34
	(3197)	(2250)	(2512)	(2548)
2/3/4	66.42	63.20	54.82	45.48
	(19690)	(7977)	(6352)	(4345)
5/6	72.99	68.06	58.75	46.49
	(7749)	(2897)	(2409)	(1612)
Total	67.49	63.76	55.54	45.93

Source: Börsch-Supan, (2022). Calculation from SHARE waves 4,5 and 6.

As an indicator of education, the SHARE survey uses the Isced (International Standard Classification of Education) code. Table 3 aggregates two or three Isced codes together to make them easier to interpret and to avoid codes with low numerosity leading to anomalous results. Analyzing the table 3, it is possible to see that for all age ranges the most common levels of education are the codes 2,3 and 4 and in particular the upper secondary education (Isced code 3) and also that the proportion of people with no or primary education increases as age increases: while in the first range they are only about 10% percent of the sample, in the range containing people over 75, about 30% percent of the sample has at maximum the primary education. This result was expected because the importance of school education has become increasingly greater as the years have passed. Regarding the responses on survival probability it is possible to observe that as education increases, it is higher within all age ranges but it is important to state that between survival probability and education there is a positive correlation that decreases with increasing age: in fact, while in the first range the difference between the survival probability of a person with no education and a person with the highest education is about 12 percentage points, in the last age range this difference is smaller than 1 percentage point.

The second characteristic that will be analyzed is the health status of the respondents. As an indicator for health the survey reports the “Sphus” variable (Self-perceived health US version)

which is an indicator defined as the subjective assessment of individual health status and contains 5 categories: Excellent, very good, good, fair and poor.

Table 4: Answer to question Ex 009 by health.

Sphus	(50-64)	(65-69)	(70-74)	(75-80)
Excellent	81.71	78.94	73.78	66.46
	(2787)	(873)	(551)	(371)
Very good	75.97	73.31	67.03	56.51
	(6652)	(2321)	(1675)	(1090)
Good	68.74	65.79	58.74	49.64
	(11892)	(5069)	(4061)	(3421)
Fair	59.55	57.68	50.41	41.35
	(7473)	(3839)	(3808)	(3802)
Poor	45.05	41.96	36.24	28.11
	(2236)	(1022)	(1178)	(1578)
Total	67.54	63.76	55.54	44.59

Source: Börsch-Supan,(2022). Calculation from SHARE waves 4,5 and 6.

From table 4 is possible to see that as expected self-perceived health has a positive correlation with survival probability, and this trend is confirmed for all 4 age ranges. It is also noteworthy that people often do not make extreme evaluations when it comes to describing their health status: in all 4 ranges, in fact, the most common responses are good and fair, while few people answered that they are in excellent or poor health. It was to be expected that health status would be positively correlated with survival probability, but it is interesting to see how large its impact is within the same age range: people with "excellent" health have a 35pp higher survival expectation than people within the same age range with "poor" health status.

The third socio-economic characteristic that is analyzed is the employment situation of respondents, and specifically the SHARE survey identifies six categories within the current job situation (cjs) variable: retired, employed, unemployed, permanently sick, homemaker and other. In table 2 I present descriptive statistics for a dummy variable that aggregates employed self-employed categories of the cjs variable.

Table 5: Answers to question Ex 009 by current job situation.

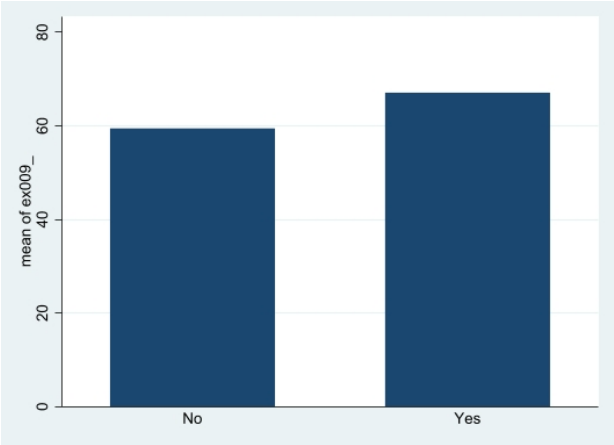
Cjs	(50-64)	(65-69)	(70-74)	(75-80)
Retired	65.57	63.57	55.53	44-66
	(8498)	(11026)	(9976)	(9007)
Employed	70.83	70.73	61.38	51.08
	(16034)	(1010)	(385)	(142)
Unemployed	64.40	-	-	-
	(16304)	(34)	(2)	(3)
Perm. Sick	53.29	51.36	38.48	34.48
	(1911)	(110)	(56)	(52)
Homemaker	65.05	59.63	54.75	44.50
	(2187)	(799)	(723)	(864)
Other	70.56	66.15	57.32	41.74
	(382)	(111)	(81)	(105)
Total	67.54	63.76	55.54	44.59

Source: Börsch-Supan, (2022). Calculation from SHARE waves 4,5 and 6.

Before analyzing Table 5, it is fair to specify that all observations with the response "not applicable" were removed from the table, which refers to all people in the nursing home who were not deemed suitable to answer questions about the work situation. The “other” category contains all people in situations different from those listed in the variable categories.

The table does not show data for unemployed people for the last three age ranges ad this is a residual category at that range with very low numerosity. Employed and self-employed people have higher survival expectation than other respondents in all 4 ranges. Considering that the survey only considers people over the age of 50, it is possible to see that apart for the first range (50-64 years) where the majority of respondents are still active in the labour market, in all the other three ranges the majority of respondents are retired. In general, it is possible to conclude that the current job situation (except for the permanently sick who are unable to work) does not have a great impact on survival probability since the average values are all quite close within the range regardless of the job situation.

Graph 3: Mean of Ex 009 over financial investments.



Source: Börsch-Supan, A. (2022). Calculation from SHARE waves 4,5 and 6.

Chart three analyzes the average of responses to question Ex 009 by whether or not they participated in the financial market. Financial market participation in this case was derived from an ad hoc created dummy variable that takes value 1 if the individual invested in bonds, stocks or mutual funds. It shows that regardless of how risky the investment is in the financial market, people who invest have a higher subjective survival expectation (67.09%) than people who decide not to invest their money (59.45%).

The last characteristic observed in the analysis of subjective survival expectation is risk attitude. This variable was derived from the response to the specific question in the SHARE survey, "Which of the statements on the card comes closest to the amount of financial risk that you are willing to take when you save or make an investment?"

Table 6: Answers to the question Ex 009 by risk attitude.

Risk	(50-64)	(65-69)	(70-74)	(75-80)
Above average	69.51	67.34	59.88	53.54
	(906)	(276)	(179)	(93)
Average	70.66	65.11	58.82	47.31
	(4927)	(1612)	(1179)	(763)
No risk	64.68	61.66	53.41	43.32
	(14886)	(7677)	(6613)	(5479)
Total	66.32	62.59	54.01	44.03

Source: Börsch-Supan, A. (2022). Calculation from SHARE waves 4,5 and 6.

As it is possible to see from table 6, most people are not risk lovers. On the contrary, many of them responded by saying that they would not take any risk within the financial market. This figure is definitely due to the fact that people over 50 have probably lower financial education than the younger generation and also the fact that they probably do not have much interest in investing their assets and risking them considering their advanced age. Survival expectation is on average higher for people who are more risk-lovers as graph 3 showed, and it is interesting to note that while in the first three ranges different levels of risk attitude do not correspond to large heterogeneity in individual's survival expectation (5pp changes), the heterogeneity becomes larger in the case of the last age range, i.e for people aged over 75: those who are more risk lovers have a subjective survival probability that is 10 percentage points higher on average than those who are risk-averse.

VI EXPECTATIONS QUESTIONS AND ALGORITHM DEVELOPMENT

After focusing on survival probability and how it correlates with various factors, it is time to get into the hearth of the thesis work and address the main topic: expectations.

The focus of this section is on the questions on subjective probabilities within the SHARE survey, and starting from those I try to understand how people form their expectations, whether they do so consciously, and whether it is possible to find patterns or common features among the responses.

Specifically, I refer to questions about expectations in which individuals are asked to estimate a probability already defined in the section III: Ex 007, Ex 008, Ex 009 and Ex 025.

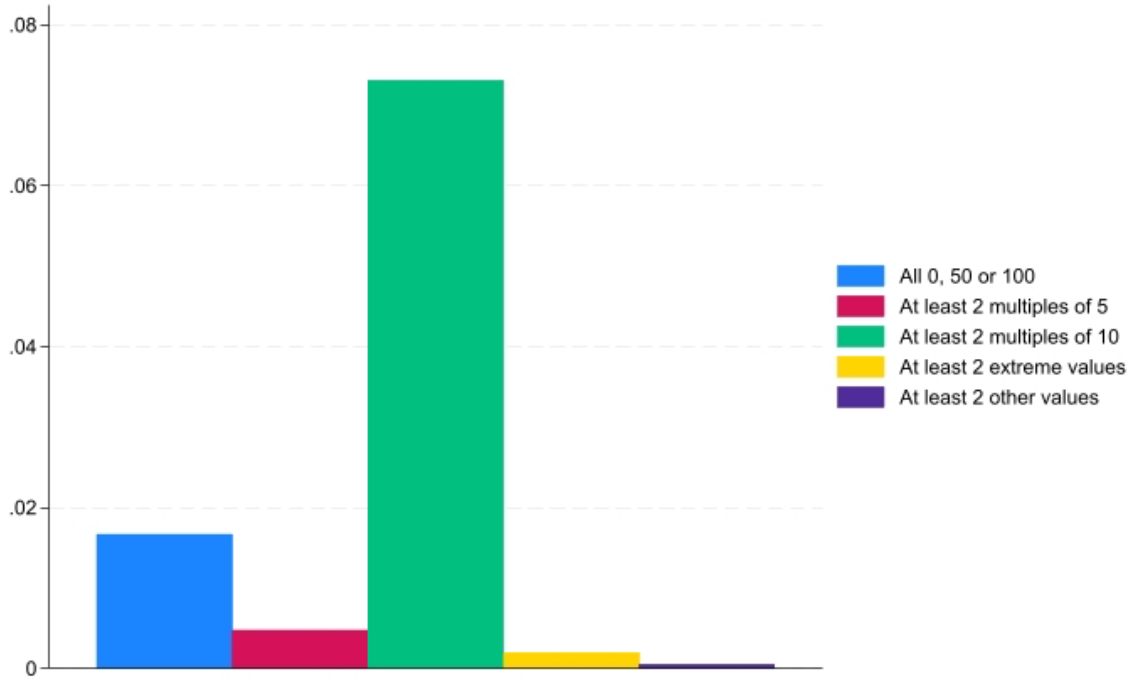
In the questions that are considered, each individual is asked to give answers that indicate a subjective probability (so a number from 0 to 100 where 0 means the individual considers the event impossible and 100 means the individual considers the event certain) about questions concerning their future, near and far. In the data description section, several descriptive statistics have proved that individuals on average seems to be able to evaluate the probability of future events. I now focus on whether the response patterns observed in previous works can be seen also in SHARE data. SHARE data also offers the opportunity to investigate whether response patterns differ by country.

The patterns that I expect to find are the use of multiples of 5 or 10, or the use of only 3 values when answering questions regarding expectations: 0,50,100.

Following Giustinelli, Manski and Molinari (2022) I am going to focus on the following patterns:

1. All 0,50,100: This group includes all respondents who always use the three values 0,50 and 100.
2. Multiples of 10: Includes all respondents who at least twice have used a multiple of 10, the values 50 and 100 (i.e 10,20,30,40,60,70,80,90)
3. Multiples of 5: Includes all respondents who at least twice have used a multiple of 5 but not 10 (i.e 5,15,25,35,45,55,65,75,85,95)
4. Extreme values: Includes all respondents who at least twice have used extreme values of the range between 0 and 100 and in particular values from 1 to 4 and from 96 to 99.
5. Other values: Includes all respondents who at least twice have used all values that were not considered by the other groups so values from 5 to 95 excluding multiples of 5 and 10.

Graph 4: Patterns in individuals' responses to the expectation questions in the SHARE survey



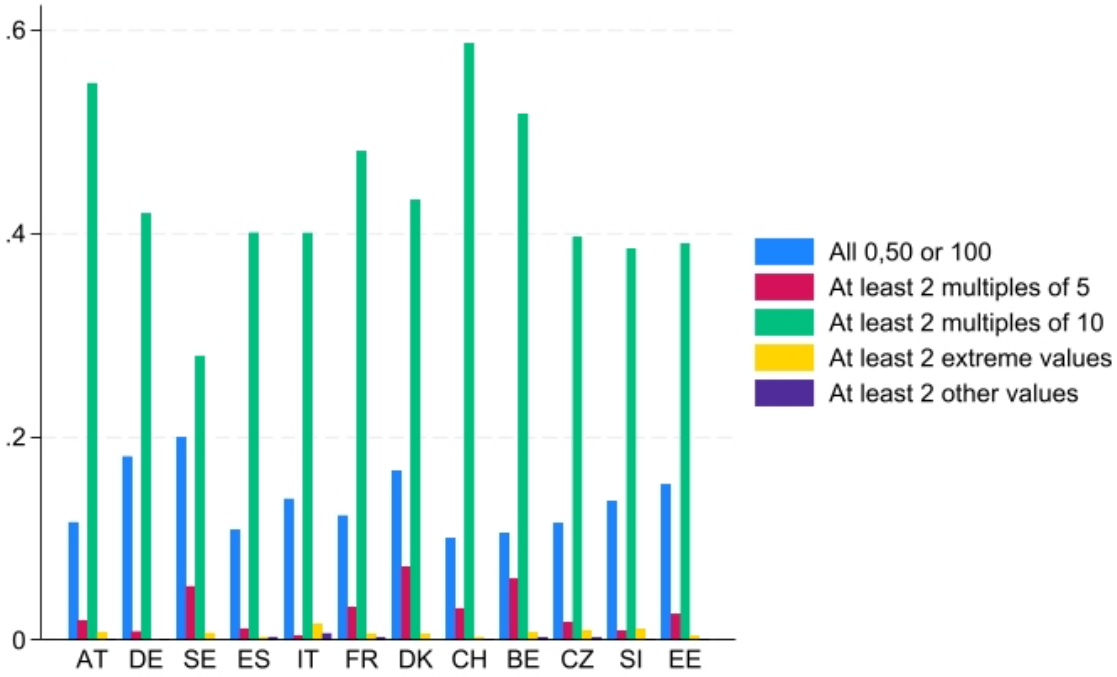
As can be seen from chart 4 and as I expected given the results presented in the literature it is possible to see that many individuals use multiples of 10 when answering questions about expectations.

There are also a large number of individuals who always use the same 3 values: 0,50 and 100 where 0 indicates that they think the event is impossible, 100 that they think it is certain and 50 that they cannot get an idea about the event. In particular, most individuals of in this category use the value 50 and this phenomenon has been explained in detail by de Bruin et al. (2000) who have explained that people use this value because it allows them to answer the question without feeling committed to a specific answer, and in their research they also showed that this phenomenon mostly affects people with lower education and older people.

Regressing the dummy variable that is 1 if individuals use only 0,50 and 100 on age and education (Isced codes) I found the same results suggested by the paper.

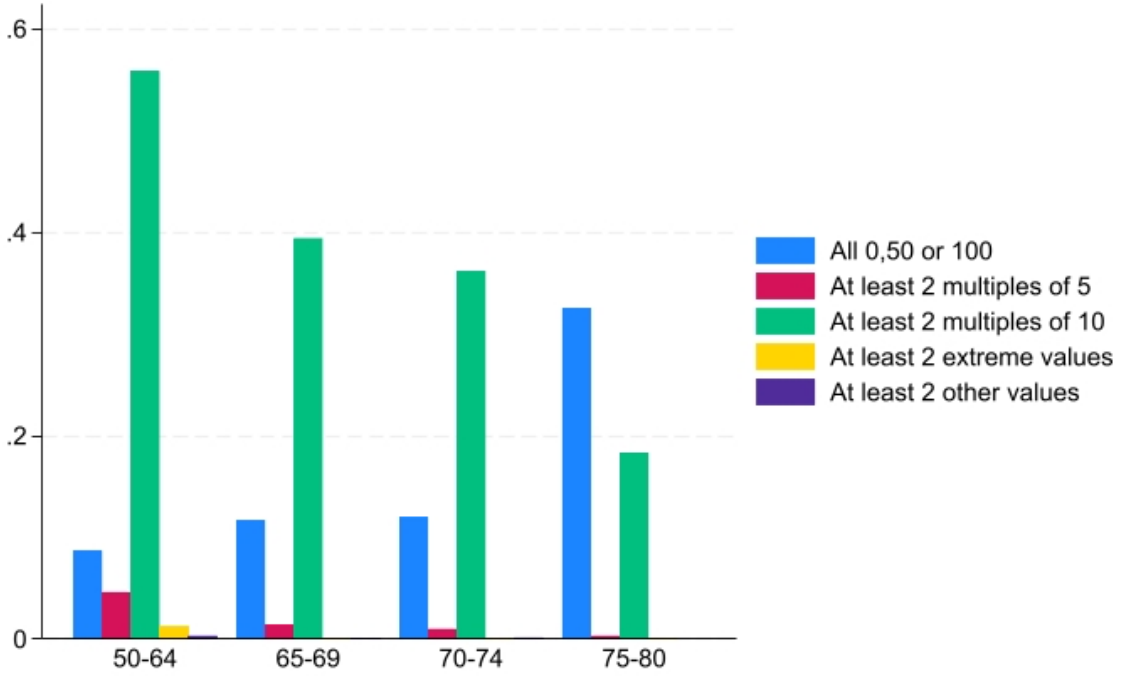
It is important to note that only this category (All 0, 50 or 100) is mutually exclusive and excludes the others while all other categories are not mutually exclusive and therefore it is interesting to understand if and how they are related to each other.

Graph 5: Patterns divided by country



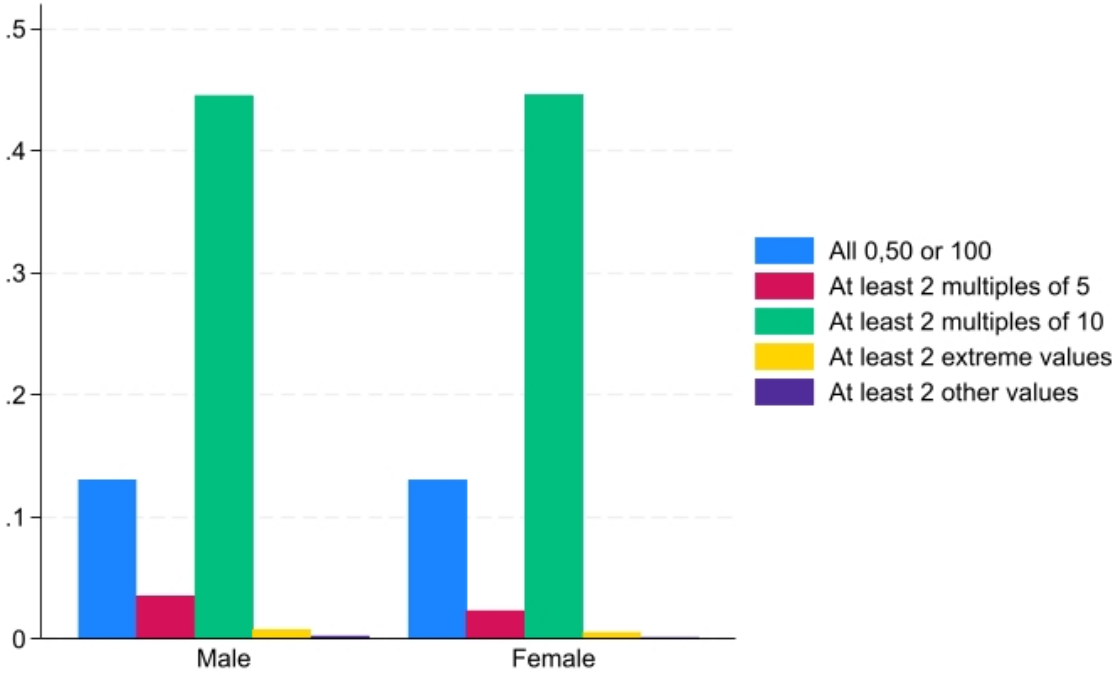
It can be seen from Figure 5 that the results in Figure 4 hold true for almost all countries. from the cross-country comparison, in fact, it can be seen that in every country the most frequently used answers to the expectation questions are multiples of 10, followed by the three focal values 0,50 and 100 and multiples of 5. Few people, on the other hand resort to extreme values or other values. The only two countries for which it is possible to see a somewhat different trend are Italy and Switzerland where people resort more often to extreme values rather than multiples of 5. In Italy, even, it is possible to see that other values (values 5 to 95 excluding multiples of 5 and 10) are also more frequently used than multiples of 5.

Graph 6: Patterns divided by age groups.



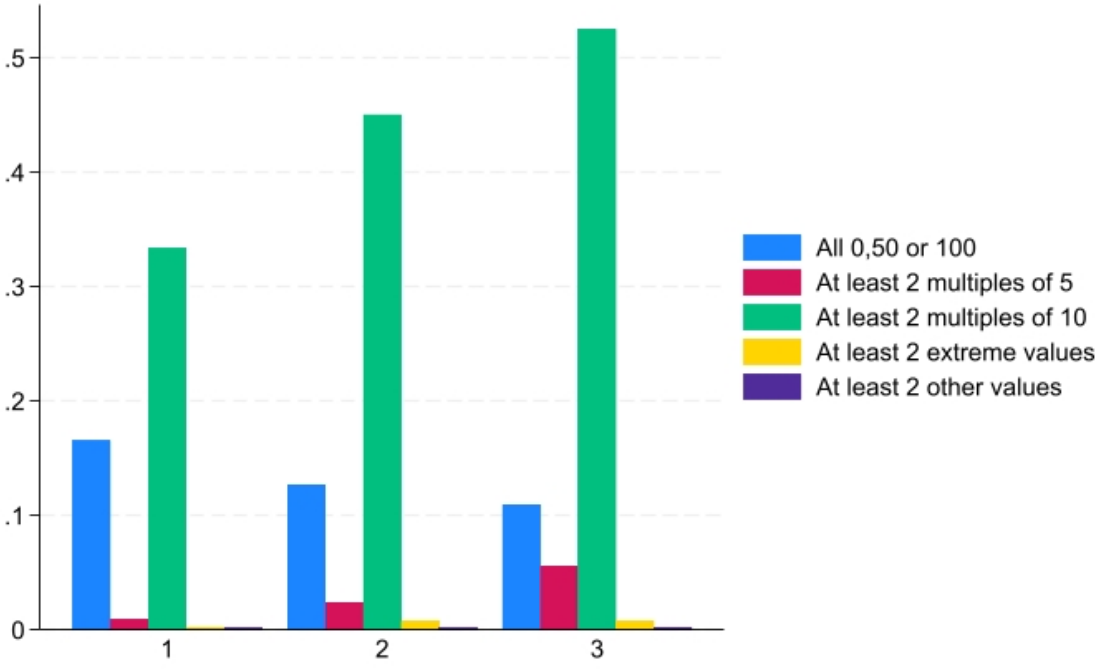
After doing a cross-country analysis of the response patterns to the SHARE survey I observe the age division. Specifically, individuals are divided into 4 groups based on the target age of the Ex 009 question. So in the first group are people aged 50 to 64 with target age 75, in the second group are individuals aged 65 to 69 with target age 80, in the third group are people aged 70 to 74 with target age 85, and finally in the last group are people over 75 with target age 90. From graph 6 it can be seen that for the first three age ranges the trend seen in graphs 4 and 5 is repeated. Here again in fact multiples of 10 are the most used, followed by the three focal values 0,50 and 100 and then by multiples of 5, extreme values and other values. Graph 6, however, shows that as people get older they tend to use fewer and fewer multiples of 5, extreme values and other values, using only multiples of 10 or 0.50 and 100. In the last range it is possible to see how people even use focal values 0.50 and 100 more often than multiples of 10. In this range the other three patterns are practically absent.

Graph 7: Patterns divided by gender



By analyzing graph 7 it is possible to see that there are no major differences between the response patterns of men and women. In both cases, as in the other graphs it is possible to find the usual trend: people prefer to use in order: multiples of 10, focal values 0,50 and 100, multiples of 5, extreme values and other values regardless of gender.

Graph 8: Patterns divided by education



For the graph 8 three education categories are distinguished: 1 (Isced codes 0,1,2), 2 (Isced codes 3,4) and 3 (Isced codes 5,6). In particular, the graph 8 shows that individuals with higher education tend to use the three focal values 0.50 and 100 less than people with lower education and prefer , instead, to use multiples of 5 more often. It also seems, that the use of extreme values and other values increases with increasing education as does the use of multiples of 5 and 10. It decreases dramatically, however, the use of 0, 50 and 100.

Graph 9: Patterns divided by couple

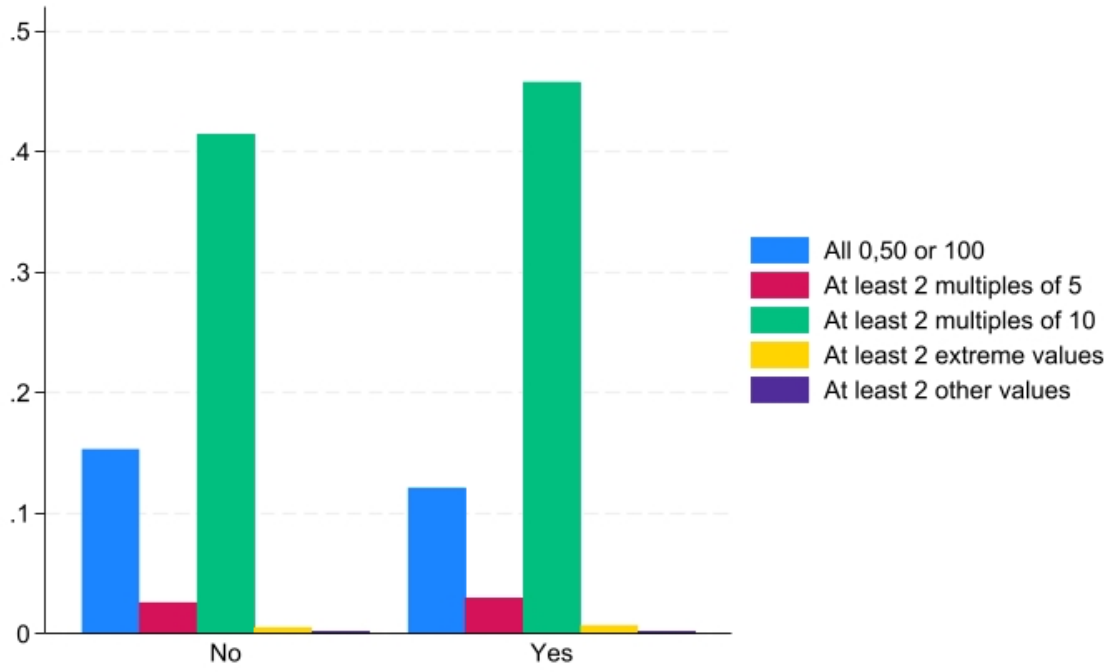


Figure 9 compares the patterns of people who are part of a couple and singles. This graph also shows the same trend as the previous ones. The most frequently used patterns are multiples of 10, the three focal values 0,50 and 100, and multiples of 5. Figure 9 also shows us that people within a couple tend to use the three focal values 0,50 and 100 less frequently, preferring to use multiples of 5, extreme values and other values.

Graph 10: patterns divided by current job situation.

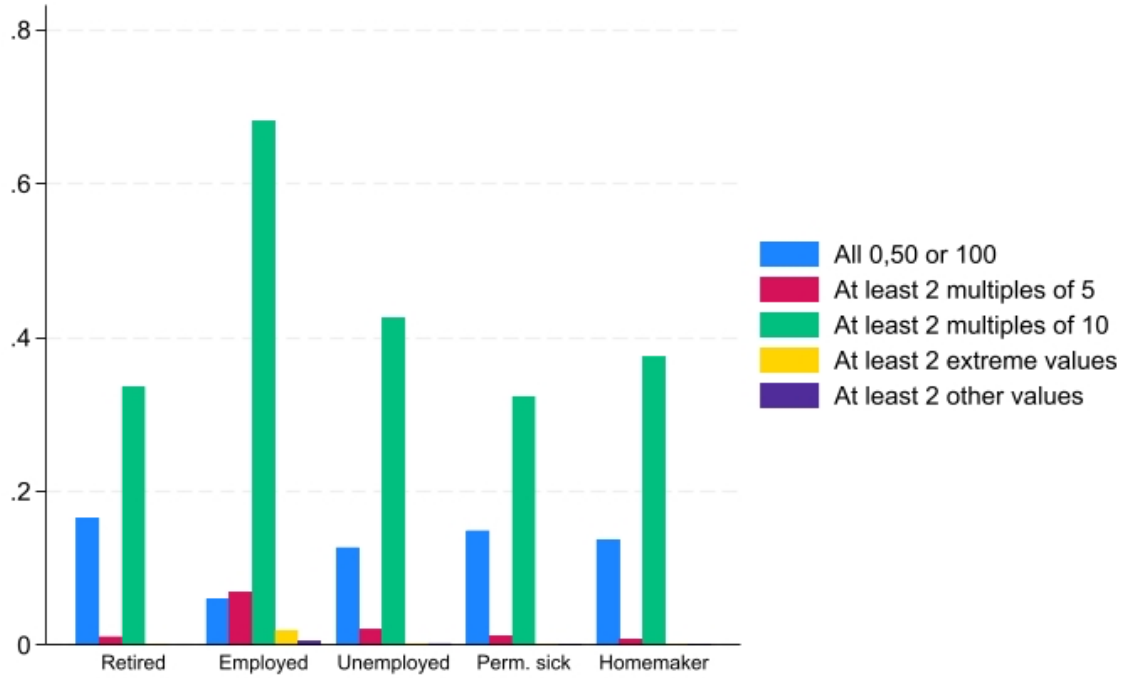


Figure 10 analyzes the differences between the patterns found within the responses to the expectations questions by dividing the respondents according to their work situation. In particular, 5 categories are distinguished: retired, employed (which also includes self-employed), permanently sick (and therefore unable to work) unemployed (regardless of whether they are looking for work or not) and homemaker. The graph shows the same trend as all others seen above for 4 out of 5 categories. The category that is an exception is the employed category. This table, as table 4 showed, is populated mainly by people between the ages of 50 and 64. In fact, from graph 10 it is possible to see that the employed is the category that differentiates its responses the most. In particular, they use multiples of 5 more often than the three focal values 0,50 and 100, and above all they use extreme values and other values much more often than the other 4 categories.

Graph 11: patterns divided by income quartiles.

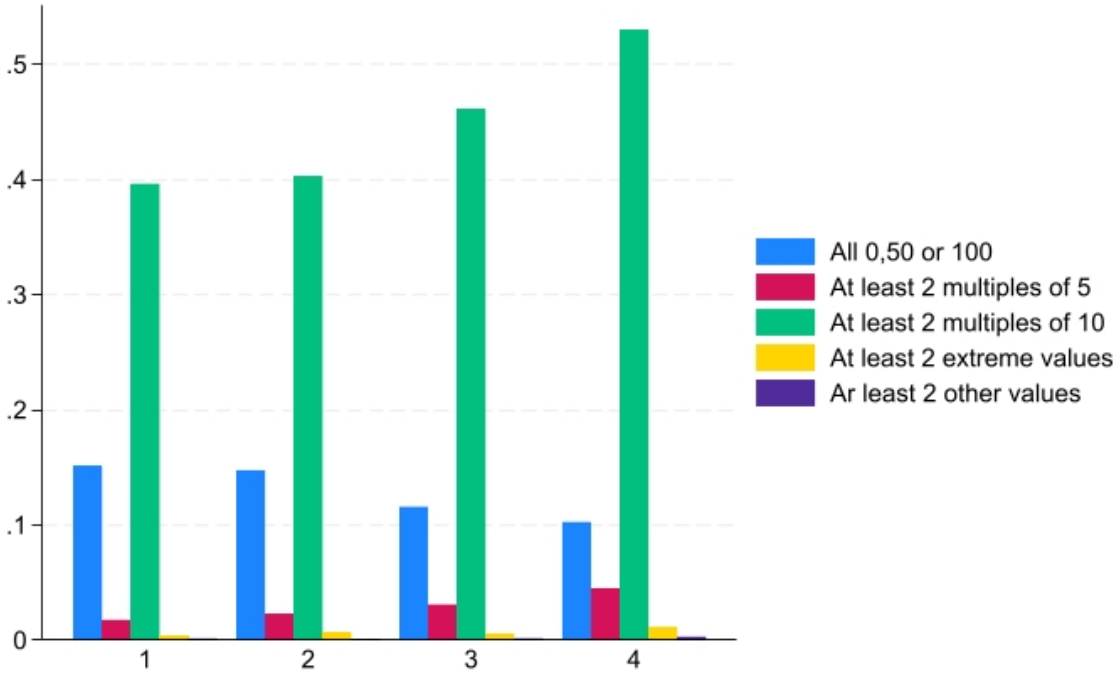


Figure 11 analyzes the difference between patterns based on total income. Four income quartiles were created for each country. From the graph, it can be seen that as wealth increases, there is a tendency to use the three focal values 0,50 and 100 less and less and instead use multiples of 5, extreme values and other values more and more. This trend is clearly visible and increases progressively as wealth increases.

Graphs 4 through 11 all showed the same trend. Although more socioeconomic control variables are included and a cross-country comparison was made. People tend to approximate their responses a lot. In general, the multiples of 10 closest to the "true" value are the ones most used, followed by the three focal values 0,50 and 100 and the multiples of 5. The percentage of people who do not round up their answers is very small and most of them tend to use extreme values (values from 1 to 4 and 96 to 100) anyway, while very few individuals use other values, i.e. those that indicate the 'true' value they have in mind.

Our results are in line with those obtained by Giustinelli et al. (2022) and Mansky and Molinari (2010). In particular, the paper by Mansky and Molinari (2010) is one of those from which Giustinelli takes his cue for the introduction of the algorithm. Specifically, in analyzing data from the Health and Retirement study they found that multiples of 10 are used at least

twice as frequently as multiples of 5, and that multiples of 50 (50 and 100) are almost always more frequent than multiples of 5. With reference to the people who tend to round more, the papers tells that it is usually people with lower education and lower income.

This analysis allows to realise that it is very difficult to use Mansky's (2004) approach that I mentioned in the introduction. That approach in fact takes as its starting point the subjective probabilities of individuals who, however, struggle to express the 'correct' value. It is therefore necessary, if this approach is to be followed, and, in general, if survey data are to be used, to check the magnitude of the measurement error and whether empirical results change when measurement error is accounted for.

Giustinelli, Mansky and Molinari (2022) proposed a solution to this problem consisting of substituting each individual point responses into interval responses calculated by an algorithm. The essential assumption of this model is the stability of responses between different questions and waves.

The ranges are constructed through an algorithm that analyzes the responses of each individual and based on the values it gives to the various responses finds for each individual a certain rounding pattern. This rounding pattern is used to create a kind of ad hoc rule to generate an interval from the point responses.

The first distinction that is made in the analysis of the responses is between values in the middle and values in the tails. Specifically by values in the middle I mean values from 25 to 75 while values in the tails are values from 1 to 24 and from 76 to 100.

After this first major distinction, the algorithm checks whether the individual uses the values 25, 50, 75 or 100 for each response and finally checks whether there are multiples of 10 or 5 within the individual's pattern of responses.

Through specific combinations that will be presented later along with the results of the algorithm adapted to the SHARE data used in this thesis the ranges of values are then created.

Specifically, the categories that I use within the algorithm to create the intervals are as follows:

-M5_t: Multiples of 5 but not of 10 in the tails: 5,15,85,95

-M10_t: Multiples of 10 in the tails: 10,20,80,90

-M100: The specific value 100

-M1_t: All values in the tails that are not multiples of 5 and10 and are not 100

-M5_c: Multiples of 5 but not of 10 in the center: 35,45,55,65

-M10_c: Multiples of 10 in the center: 30,40,60,70

M25: The values 25 and 75

M50: The specific value 50

M1_c: All values in the center that are not multiples of 5 and 10 and are not 25.50 or 75

Table 8: Combinations to form the intervals within the algorithm.

	M1_t	M5_t	M10_t	M100	No answers in t
M1_c	[x,x]	[x,x]	[x,x]	[x,x]	[x,x
M5_c	[x,x]	[Max (0,x-2), Min (100, x+2)]	[Max (0,x-2), Min (100, x+2)]	[Max (0,x-2), Min (100, x+2)]	[Max (0,x-2), Min (100, x+2)]
M10_c	[x,x]	[max (0,x-2), Min (100, x+2)]	[max (0,x-5), Min (100, x+5)]	[Max (0,x-5), Min (100, x+5)]	[Max (0,x-5), Min (100, x+5)]
M25	[x,x]	[max (0,x-2), Min (100, x+2)]	[max (0,x-5), Min (100, x+5)]	[Max (0,x-12), Min (100, x+12)]	[Max (0,x-12), Min (100, x+12)]
M50	[x,x	[max (0,x-2), Min (100, x+2)]	[max (0,x-5), Min (100, x+5)]	[Max (0,x-25), Min (100, x+25)]	[Max (0,x-25), Min (100, x+25)]
No answers in c	[x,x]	[max (0,x-2), Min (100, x+2)]	[max (0,x-5), Min (100, x+5)]	[Max (0,x-50), Min (100, x+50)]	[0, 100]

To better understand table 8, I need to specify that 'x' means the original value of the respondent's answer and that to create a range it is necessary that at least two answers are available for each individual because the range is formed either by the combination of two conditions or if the same condition is repeated at least twice within each individual answers.

Some expectation questions are asked to SHARE respondents only once, when they enter the sample (Ex 007, Ex 008, Ex 025). The survival expectation questions (ex009) are instead asked in each wave. Assuming stability of the response patters over time, for each individual I consider all the answers to expectation questions polling together wave 4, 5 and 6. This means that the number of questions used for each individual range from 3 to 6.

VII MATLAB CODE FOR THE ALGORITHM

In this section I report the code used on Matlab for the algorithm that forms the intervallic responses.

Individuals' responses to the expectations questions were placed in an excel file called algorithm data, a file imported into Matlab.

```
id = algorithm_data(:, 1);
data = algorithm_data(:, [2:13]);
data_array = table2array(data);
data_array = double(data_array);
center_logical = (data_array >= 25) & (data_array <= 75);
tail_logical = (data_array < 25) | (data_array > 75);
M10_t = tail_logical & mod(data_array, 10) == 0;
M5_t = tail_logical & mod(data_array, 10) == 5;
M1_t = tail_logical & ~M10_t & ~M5_t;
M100 = data_array==100;
M10_c = center_logical & mod(data_array,10)==0;
M5_c = center_logical & mod(data_array, 10) == 5;
M1_c = center_logical & ~M10_c & ~M5_c;
M25 = data_array == 25 | data_array == 75;
M50= data_array ==50;
NR = ~(data_array >= 0 & data_array <= 100);
num_rows = size(data_array, 1);
num_columns = size(data_array, 2);
interval_estimates = num2cell(data_array);
for i = 1:num_rows
    for j = 1:num_columns
        answer = data_array(i, j);
        if ~isnan(answer)
            if sum (M1_t(i, :)) >= 1 || sum(M1_c(i, :)) >= 1
                assigned_value = [answer; answer];
            elseif sum(M5_t(i, :)) >= 2 || sum(M5_t(i, :)) >= 1 &&
sum(M5_c(i, :)) >= 1 || sum(M5_t(i, :)) >= 1 && sum (M10_c (i,:)) >= 1 ||
sum(M5_t(i, :)) >= 1 && sum (M25 (i,:)) >=1 || (sum(M5_t(i, :)) >= 1 &&
sum(M50(i, :)) >= 1) || (sum(M10_t(i, :)) >= 1 && sum(M5_c(i, :)) >= 1)||
sum(M100 (i, :)) >= 1 && sum(M5_c(i, :)) >= 1 || sum (M5_c (i,:))>= 2
                assigned_values = [max(0, answer - 2); min(answer+ 2,100)];
            elseif sum(M10_t(i, :)) >= 2 || sum(M10_c(i, :)) >= 2 ||
sum(M10_t(i, :)) >= 1 && sum(M10_c (i, :)) >= 1 || sum(M10_t(i, :)) >= 1 &&
sum (M25 (i,:))>=1 || sum(M10_t(i, :)) >= 1 && sum(M50 (i, :)) >= 1 ||
sum(M100 (i, :)) >= 1 && sum(M10_c(i, :)) >= 1
```

```
                assigned_value= [max(0, answer - 5); min(answer + 5, 100)];
            elseif sum(M100(i, :)) >= 1 && sum(M25(i, :)) >= 1 ||
sum(M25(i, :)) >= 2
                assigned_value = [max(0, answer - 12.5); min(answer + 12.5,
100)];
            elseif sum(M100(i, :)) >= 1 && sum(M50(i, :)) >= 1 ||
sum(M50(i, :)) >= 2
                assigned_value = [max(0, answer - 25); min(answer + 25,
100)];
            elseif sum (M100(i,:)) >=2
                valore_assegnato= [max(0, answer - 50); min(answer + 50,
100)];
            else
                assigned_valuea = [0;100];
            end
            if ~isempty(assigne_value)
                interval_estimates{i, j} = assigned_value;
            end
        end
    end
end
```

VIII EMPIRICAL APPLICATION

In this chapter the data I have collected so far will be used to estimate a model that seeks to analyze the correlation between an indicator of subjective survival expectation and financial market participation of individuals who responded to the SHARE survey.

Specifically, the model will analyze the correlation between the response to question Ex 009 of the SHARE survey on subjective survival probability and the presence within an individual's portfolio of bonds, stocks or mutual funds.

I chose to analyze how people invest based on their subjective survival expectation because it is a very relevant topic. As mentioned in the literature review for this empirical application I borrow from the paper by Spaenjers and Spira (2016).

The goal of their paper is to understand how life expectancy, their measure of an individual's investment horizon, influences investments within the financial market. They use data taken from the Survey of Consumer Finance, which also includes questions on gender, education, health status, optimism, and many other factors.

As might be expected, the results of their paper show that investors with higher life expectancy invest more in riskier securities (stocks) than those with a lower investment horizon who opt for safer securities. The results remain significant even when factors that may influence portfolio choice, such as risk tolerance and financial goals, are taken into account.

In my thesis I am not going to analyze investment riskiness but only financial market participation. Specifically, I derived financial market participation through an indicator contained in the SHARE data called "bsmf" which indicates the total amount each individual has invested at the time of the survey in bonds, stocks or mutual funds, and from this variable I created a dummy variable with value 1 if the bsmf indicator takes positive values and value 0 if the respondent has invested 0.

The model will analyze only the financial respondent's answers: within the SHARE survey for each household, a financial respondent is identified and he/she answers all questions related to the financial sphere.

The model therefore will consider only one observation per household and the individual characteristics are those of the financial respondent. Additionally, only individuals who are between 50 and 80 years old who report valid answers to the subjective survival question (ex009) in wave 6 are part of the sample.

The final sample size is 16789.

The dependent variable of the model is financial market participation while the main independent variable of interest is subjective survival probability. The model I estimate is the following:

$$fin\_part = \beta_0 + \beta_1 subj\_survival + \boldsymbol{X'\beta_2} + \beta_3\, health + \boldsymbol{Z'\beta_4} + \boldsymbol{country\beta_5} + \boldsymbol{e} \quad (\mathbf{1})$$

In this equation the vector X includes demographic controls such as:

- Age
- Age squared
- Single
- Children: dummy variable with value 1 if the respondent has 1 or more children and 0 otherwise.

The variable health refers to a dummy with value 1 if the respondent answers poor to the question about self-perceived health and the vector country contains all the country dummies.

The vector Z includes socio-economic controls such as:

- Employed
- Income quartiles
- Secondary education: A dummy that takes value 1 if an individual's Isced code is greater than 3.
- Financial literacy: a dummy with a value of 1 if the individual's numeracy score is higher than 2.

All variables not defined here are the same as those already defined in Table 2.

I estimate the model with OLS, I therefore estimate a linear probability model, results are reported in tables 9 and 10, robust standard errors are reported in parentheses.

The model will first consider the point responses provided by individuals to the question Ex 009 in the sixth wave and then the estimates that account for the interval responses calculated through the algorithm I presented in the previous section so as to see if the way the individual rounds up his or her response affects the result. To do that I estimate the model in equation (1) following the multiple imputation technique proposed by Little and Rubin (1987) to formally account for the variability within and between the sets of imputations produced.[1]

To account for measurement error, and therefore to use within the model the intervals for subjective survival probability, I generate 10 multiple imputations for the subjective survival probability. I do that by taking 10 random draws from a Uniform distribution in the interval

[1] I used the `mi import` and `mi estimate` commands on Stata 18.0 that are part of the `mi package` and allow us to estimate models with data having multiple imputations for missing values.

(a,b), where a and b are the extremes of the interval generated by the algorithm for each individual.

Table 9: Regression of financial investments on subjective survival probability divided by age, point estimates.

	50-80	50-65	66-80
Subj. survival probability	.00039*** (.00009)	.00073*** (.00015)	.00016 (.00012)
Age	.02543*** (.00745)	.08027* (.004197)	.08543** (.03375)
Age Squared	-.00017*** (.00006)	-.00062* (.00035)	-.00058** (.00023)
Reference country: Germany			
Austria	-.07990*** (.01969)	-.04318 (.03242)	-.09978*** (.02510)
Sweden	.39460*** (.02475)	.35753*** (.04739)	.38934*** (.02974)
Spain	-.15077*** (.019503)	-.12669*** (.03204)	-.16425*** (.02497)
Italy	-.02732 (.02120)	.02609 (.03606)	-.05269** (.02656)
France	-.03191 (.02011)	.00452 (.03289)	-.05408** (.02593)
Denmark	.16570*** (.02362)	.17830*** (.03693)	.15761** (.03168)
Switzerland	.14682*** (.02133)	.11924*** (.03421)	.17452*** (.02783)
Belgium	.10903*** (.02052)	.14277*** (.03308)	.09443*** (.02685)
Czech Republic	-.11030*** (.01911)	-.05919* (.03213)	-.13706*** (.02404)
Slovenia	-.11908*** (.02023)	-.06483* (.03337)	-.15371*** (.02585)

Estonia	-.19267*** (.01827)	-.16465*** (.03046)	-.20784*** (.02314)
Woman	-.04315*** (.00634)	-.02922** (.00966)	-.05088*** (.00861)
Secondary Education	.08735*** (.00738)	.07733** (.01113)	.09408*** (.01001)
Single	-.00498 (.00746)	-.01195 (.01128)	.00559 (.01030)
Children	-.05296*** (.01060)	-.07914*** (.01624)	-.03105** (.01417)
Poor Health	-.03729*** (.00797)	-.03194** (.01402)	-.03402*** (.00995)
Employed	-.00448 (.00953)	.01389 (.01109)	-.01760 (.02287)
Reference category: 1st income quartile			
2nd Income Quartile	.05892*** (.00766)	.05612*** (.01284)	.05490*** (.00973)
3rd Income Quartile	.09472*** (.00907)	.08208*** (.01368)	.10393*** (.01231)
4th Income Quartile	.18962*** (.01024)	.16902*** (.01460)	.20343*** (.01433)
Financial literacy	.03956*** (.00885)	.012618 (01716)	.04767*** (.01052)
Constant	-0.80502*** (0.25277)	-2.44969** (1.24367)	-2.96836** (1.22645)
N	16789	7187	9602
R_SQUARED	0,2086	0,1785	0,2418

Note: Significant at 10% level (*), 5% level (**), 1% level (***). Standard errors robust to heteroskedasticity in parentheses.

Table 10: Regression of financial investments on subjective survival probability divided by age, interval estimates.

	50-80	50-65	66-80
Subj. survival probability	.00031**	.00071***	.00012
	(.00012)	(.00020)	(.00015)
Age	.02930***	.05878	.09529**
	(.00808)	(.04329)	(.03510)
Age Squared	-.00020***	-.00044	-.00065**
	(.00006)	(.00036)	(.00024)
Reference country: Germany			
Austria	-.07683***	-.04489	-.09423***
	(.01990)	(.03248)	(.02532)
Sweden	.39402***	.36182***	.39053***
	(.02511)	(.04767)	(.02999)
Spain	-.15039***	-.12314***	-.16842***
	(.01997)	(.03277)	(.02526)
Italy	-.02657	-.02514	-.05003
	(.02150)	(.03634)	(.02677)
France	-.02832	-.00442	-.04977
	(.02038)	(.03298)	(.02616)
Denmark	.16333** *	.17927***	.15693***
	(.02384)	(.03702)	(.03188)
Switzerland	.14731***	.12053***	.17874***
	(.02157)	(.03432)	(.02801)
Belgium	.11459***	.14512***	.09866***
	(.02079)	(.03322)	(.02711)
Czech Republic	-.10710***	-.06066*	-.13228***
	(.01951)	(.03243)	(.02444)
Slovenia	-.11435***	-.06258*	-.15245***
	(.02058)	(.03362)	(.02607)
Estonia	-.19087***	-.16594**	-.20532***
	(.01847)	(.03058)	(.02329)
Woman	-.04139***	-.02883***	-.04876***
	(.00661)	(.00997)	(.00893)

Secondary Education	.08893*** (.007630)	.07646*** (.01136)	.09746*** (.01034)
Single	-.00523 (.00777)	-.01086 (.01162)	.00462 (.01063)
Children	-.05457*** (.01108)	-.07982*** (.01675)	-.03253*** (.01475)
Poor Health	-.04212*** (.00863)	-.03771 (.01491)	-.03863*** (.01058)
Employed	-.00639 (.00985)	.01121 (.01146)	-.02067 (.02355)
Reference category: 1st income quartile			
2nd Income Quartile	.06473*** (.00801)	.06035*** (.01328)	.05770*** (.01016)
3rd Income Quartile	.09825*** (.00945)	.08590*** (.01413)	.10532*** (.01268)
4th Income Quartile	.19575*** (.01064)	.17511*** (.01509)	.20552*** (.01471)
Financial literacy	.04363*** (.00967)	.01928 (.01900)	.04892*** (.01115)
Constant	-0.92713*** (0.27268)	-1.81112*** (1.28302)	-2.1964*** (1.2345)
N	16789	7187	9602

Note: Significant at 10% level (*), 5% level (**), 1% level (***). Standard errors robust to heteroskedasticity in parentheses.

The first column of table 9 suggests that there is a positive and significant correlation between subjective survival expectation and the investment in the financial market. It is important to notice that my measure of life horizon, the subjective survival probability of reaching a certain target age, is less intuitive than a measure of life expectancy. To assess the magnitude of the parameter estimates, I compute the change in the probability of investment in the financial market driven by a 25 percentage points increase in the probability of surviving to

the set target age. The parameter estimates for the entire sample implies an increase in the probability of investment in the financial market of about 1 percentage point.

Age, has a positive effect on financial market participation up to an 70 and 75 and at older ages financial market participation reduces. This is confirmed by the estimates in column 2 and 3 of table 9 that report results for the same model estimated separately for individuals aged 50-65, or older.

As can be seen in Table 9, Germany was taken as the reference group for the cross-country comparison. The results as far as countries are concerned correspond to what could be expected. It is possible to see, in fact, that in Northern European countries I find larger participation in the financial market, while in the cases of Southern and Eastern European countries there is a parameter, suggesting a lower participation compared to Germany. In particular, from the first column of Table 9 it is possible to observe that the country with the highest participation is Sweden: in fact, being Swedish increases the probability of investing within the financial market by 39pp compared to living in Germany. On the other hand, the country with the lowest participation is Estonia. Living in Estonia, in fact, makes one less likely to participate in the financial market by 20pp, compared to Germany.

According to the model estimates, being female decreases the probability of investing within the financial market by 4 percentage points.

Education has a positive effect on financial market participation. In particular, Table 9 shows that having an Isced code greater than 3 increases the probability of investing within the financial market by 8.8 percentage points.

Being single decreases the probability of investing within the financial market, but from Table 9 it can be seen that the variable is not significant.

In contrast, the children's variable is significant and gives a surprising result. In the literature, in fact, it is possible to find many papers that say that bequest motives are a reason for people to invest more. For example, the paper by De Nardi and Yang (2014) says that households driven by bequest motives invest more by having a longer time horizon and also shows that these households tend to invest in riskier securities than others. The results in the first column of Table 9, however, show that having children, in this case, reduces the probability of participating in the financial market by 5,5 percentage points.

The poor health variable, as expected, has a negative impact on financial market participation (-4.2pp).

After the demographic variables, I analyze the effect of the economic variables. Specifically, the first control variable is based on current job situation, being employed (or self-employed) does not significantly correlate with financial market participation.

Another variable analyzed is income which is found to have a very strong correlation with the probability of investing in the financial market. Specifically, while between the first and second quartiles the difference is about 6pp and between the second and third quartiles the difference is 4pp, the big difference is between the fourth quartile and all others. Those in the fourth quartile are in fact about 19pp more likely to invest than those in the first quartile, 14pp more likely than those in the second quartile, and 9pp more likely than those in the third.
The last variable analyzed is financial literacy. This variable also gives a predictable result. In fact, when an individual reports a value greater than 3 on the financial literacy questions in the SHARE survey, his or her probability of participating within the financial market increases by 4.3pp.
After the analysis of the general results obtained in the first column of Table 9 that contained the entire sample in terms of age, it is now interesting to see how the correlations found by the model change if I split the entire sample into two parts: 50-65 and 66-80.
Starting with the subjective survival probability, in the second column of Table 9 it is possible to see that the correlation with financial market participation is higher than in the first column. Indeed, in this case, a 25 percent increase in an individual's survival probability makes him or her more likely to participate in the financial market by 1.8pp.
The opposite effect, on the other hand, can be seen in the third column. In this case the correlation is much lower and even not statistically significant.
The time horizon matters more for the younger individuals considered than for the older.
Analyzing the difference between the three columns related to the country rows, it is possible to see that the trend is the same for almost all countries. In fact, Table 9 shows that whether or not a person belongs to a country has more "influence" in a person's investment choice if that person is older (65-80). The magnitude of the difference between the two age ranges increases if the effect is negative. This means that if the correlation between living in that country and financial market participation is negative, this impact will be much more pronounced in people aged 65-80. The only country where a different trend can be seen is Belgium. In fact, in Belgium, there are people in the first age range (50-65) who have a higher correlation with the probability of financial market participation.
Continuing down in Table 9, I see that the negative effect of being female on the probability of participating within the financial market increases as age increases. In fact, a woman between the ages of 50 and 65 has a negative correlation of 2.8pp, which instead becomes 4.8pp if she is between 66 and 80 years old. The same argument can be made for education as well. Education seems to be a more important discriminant within the probability of financial market participation for people between 66 and 80 than for people between 50 and 65

although the effect is still quite large: 7.6 percentage points for the first age range and 9.7 percentage points for the second age range.

Single is never significant. The parameter of having children is negative and larger in magnitude for people between 50 and 65 (-7.9pp). However, the correlation between children and financial market participation remains negative, but lower, for people over 66 (-3.3pp).

The effect of poor health is more or less the same across the three columns in Table 9, the only difference being that for the second column (individuals aged 50 to 65) the effect becomes non-statistically significant.

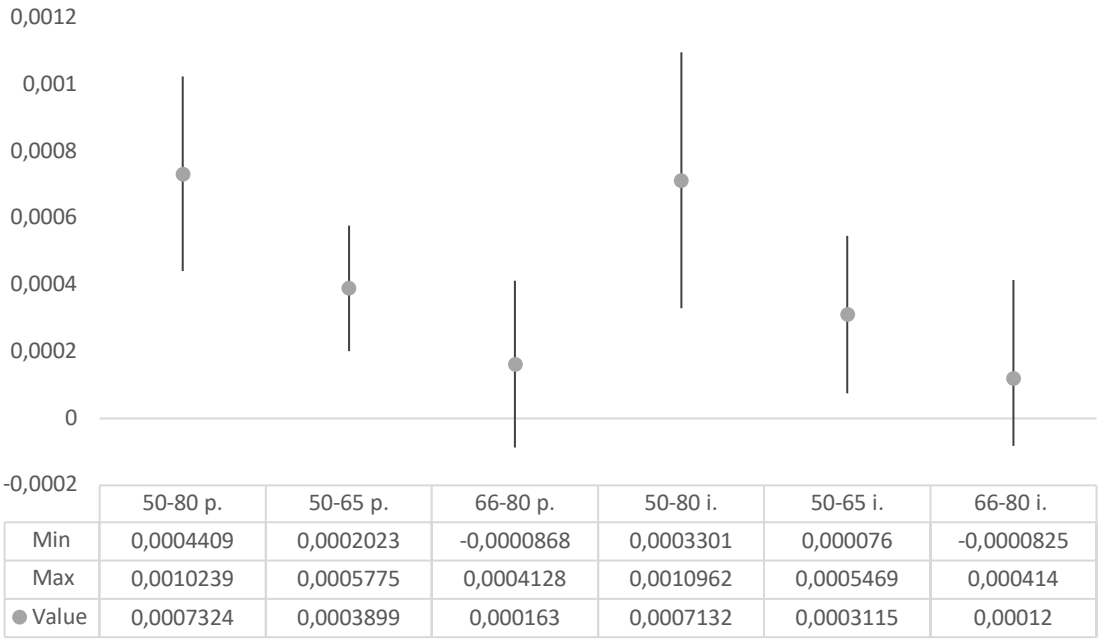
The last variable observed is financial literacy. Comparing the two age ranges allows to see how it becomes much more influential in the choice of financial market participation for people over 65 than for those age 65 or less.

After analyzing the results of Table 9, obtained using the responses given by individuals to the SHARE survey questions, I analyze the results in Table 10, which accounts for measurement error using the empirical strategy introduced at the beginning of this section. In commenting Table 10, I focus on the subjective survival probability regressor only. The other controls show results in line with those obtained in Table 9.

Analyzing the first row of the first column of Table 10 I see that the correlation between subjective survival probability and financial investment is similar to what we found before. A 25 percent increase in survival probability, in fact, increases his probability of participating within the financial market by 0.75 percentage points. Comparing this value with the value obtained in Table 9 it is possible to see that in this case the correlation found is slightly weaker and as expected has larger standard errors, however it remains statistically significant. Even accounting for measurement error, the results confirm that subjective survival have a significant effect on financial market participation.

The effect of survival probability is higher in the first age range (50-65) where a 25 percent increase in survival probability increases the probability of investment by 1.8pp. In contrast, the effect again becomes, as in Table 9, very small and statistically insignificant in the case of those aged 66-80. In both cases the standard errors increase, but this does not affect our main conclusion on the importance of subjective survival probability on financial market participation.

Graph 11: 95% confidence intervals for calculated correlations between subjective survival expectation and financial market participation.



	50-80 p.	50-65 p.	66-80 p.	50-80 i.	50-65 i.	66-80 i.
Min	0,0004409	0,0002023	-0,0000868	0,0003301	0,000076	-0,0000825
Max	0,0010239	0,0005775	0,0004128	0,0010962	0,0005469	0,000414
Value	0,0007324	0,0003899	0,000163	0,0007132	0,0003115	0,00012

The legend below chart 11 helps us understand. Age ranges followed by the letter p refer to the results obtained using point estimates (table 9). The letter i after the age range, on the other hand, indicates the results obtained in table 10. From Figure 11 it can be seen that for all three age groups considered (50-80), (50-65) and (65-80) the values obtained in Table 9 and Table 10 with the interval estimates are very similar to each other. What stands out, however, is the width of the confidence interval. Indeed, in the case of the results obtained with the interval estimates, a wider confidence interval can be seen, indicating greater variability.

IX CONCLUSIONS

This thesis started from the idea of Mansky (2004) that, in order to understand how people form their expectations, it is necessary to analyze subjective probability data mixed with other individual data (health, education, wealth).

I therefore started from the SHARE data, and show that subjective survival probability contains information to exploit: in particular they vary as expected by gender, education, income and many different characteristics. I then describe and analyze possible common response patterns among the survey participants.

This thesis demonstrated the presence of these trends through graphs that show how it is possible to identify common patterns between the responses of all people. This happens because people tend to round their responses, often responding to probability questions using easily communicated values. I show that the most common patterns are recursive but their use changes with individual characteristics, such as country of residence, gender, age, education, suggesting the presence of response styles.

The present study therefore set the goal of introducing a procedure that would give us an understanding of the magnitude of the measurement error. I then explore with an empirical application whether accounting for measurement error affects our conclusions. I followed the approach proposed by Giustinelli et al. (2022) to transform the point-based responses of an individual into interval responses that express the uncertainty contained into their answer. This was made possible by creating a rounding rule ad hoc for each respondent based on the responses to the expectations they gave in the waves examined.

The empirical application that I proposed look at the potential role of subjective survival expectations in explaining participation into the financial market. From the comparison between the two estimated models, using point responses and intervals in turn, it emerged that the difference between point estimates and interval estimates is small. As expected standard errors increase but this does not affect our conclusions. In particular a 25% increase in the subjective survival probability increases participation into the financial market by about 1pp, the effect almost doubles when the focus is on individuals aged between 50 and 65.

I can conclude that measurement error in this specific empirical application does not affect my conclusions. However, I have developed a simple strategy that allows to easily check for the role of measurement error in empirical application involving the use of subjective probability questions, provided that we have enough probability questions answered by the same individual to analyze his/her response pattern.

X REFERENCES

- Bloom, David E., et al. "The effect of subjective survival probabilities on retirement and wealth in the United States." (2006)

- Blattman, Christopher, et al. "Measuring the measurement error: A method to qualitatively validate survey data." *Journal of Development Economics* 120 (2016): 99-112.

-Börsch-Supan, A. (2022). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 4. Release version: 8.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w4.800

-Börsch-Supan, A. (2022). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 5. Release version: 8.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w5.800

-Börsch-Supan, A. (2022). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 6. Release version: 8.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w6.800

-Bound, John, Charles Brown, and Nancy Mathiowetz. "Measurement error in survey data." *Handbook of econometrics*. Vol. 5. Elsevier, 2001. 3705-3843.

-De Bruin, Wändi Bruine, et al. "Verbal and numerical expressions of probability:“It's a fifty–fifty chance”." *Organizational behavior and human decision processes* 81.1 (2000): 115-131.

- De Nardi, Mariacristina, and Fang Yang. "Bequests and heterogeneity in retirement wealth." *European Economic Review*72 (2014): 182-196.

-England, Paula. "The case for marriage: why married people are happier, healthier, and better off financially." *Contemporary Sociology* 30.6 (2001): 564.

- Giustinelli, Pamela, Charles F. Manski, and Francesca Molinari. "Tail and center rounding of probabilistic expectations in the health and retirement study." Journal of econometrics 231.1 (2022): 265-281.

- Juster, F. Thomas. "Consumer buying intentions and purchase probability: An experiment in survey design." *Journal of the American Statistical Association* 61.315 (1966): 658-696.

-Lachmann, Ludwig M. "The role of expectations in economics as a social science." *Economica* 10.37 (1943): 12-23.

- Lichtenstein, Sarah, and J. Robert Newman. "Empirical scaling of common verbal phrases associated with numerical probabilities." *Psychonomic science* 9 (1967): 563-564.

-Lusardi, Annamaria, and Olivia S. Mitchell. "Financial literacy and retirement planning: New evidence from the Rand American Life Panel." *Michigan Retirement Research Center Research Paper No. WP* 157 (2007).

- Machlup, Fritz. "Marginal analysis and empirical research." *The American Economic Review* 36.4 (1946): 519-554.

- Manski, Charles F. "Measuring expectations." *Econometrica* 72.5 (2004): 1329-1376**.**

- Manski, Charles F., and Francesca Molinari. "Rounding probabilistic expectations in surveys." *Journal of Business & Economic Statistics* 28.2 (2010): 219-231.
-Mislevy, R. J. (1991). [Review of *Statistical Analysis with Missing Data*, by D. B. Rubin & Roderick J. A. Little]. *Journal of Educational Statistics*, *16*(2), 150–155.
- Morrison, Donald G. "Purchase intentions and purchase behavior." *Journal of marketing* 43.2 (1979): 65-74.
- McClelland, Alastair GR, and Fergus Bolger. "The calibration of subjective probability: Theories and models 1980–94." (1994).
- Saris, Willem E., and Melanie Revilla. "Correction for measurement errors in survey research: necessary and possible." *Social Indicators Research* 127 (2016): 1005-1020.
- Spaenjers, Christophe, and Sven Michael Spira. "Subjective life horizon and portfolio choice." *Journal of Economic Behavior & Organization* 116 (2015): 94-106.
- van der Beck, Philippe, and Coralie Jaunin. "The equity market implications of the retail investment boom." *Swiss Finance Institute Research Paper* 21-12 (2021).
- Van der Klaauw, Wilbert. "On the use of expectations data in estimating structural dynamic choice models." *Journal of Labor Economics* 30.3 (2012): 521-554.